MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES
Instituto Brasileiro de Informação em Ciência e Tecnologia

MINISTÉRIO DA SAÚDE
Agência Nacional de Vigilância Sanitária

GUIA SOBRE A CONSTRUÇÃO DE TESAUROS

ibict
Instituto Brasileiro de Informação em Ciência e Tecnologia
UNIDADE DE PESQUISA DO MCTI

ANVISA
Agência Nacional de Vigilância Sanitária

PRESIDÊNCIA DA REPÚBLICA

Jair Messias Bolsonaro
Presidente da República

Hamilton Mourão
Vice-Presidente da República

MINISTÉRIO DA SAÚDE

Eduardo Pazuello
Ministro da Saúde

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA

Antonio Barra Torres
Diretor-presidente/Primeira Diretoria

Meiruze Sousa Freitas
Segunda Diretoria

Cristiane Rose Jourdan Gomes
Terceira Diretoria

Romison Rodrigues Mota
Quarta Diretoria

Alex Machado Campos
Quinta Diretoria

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

Marcos Cesar Pontes
Ministro da Ciência, Tecnologia e Inovações

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA

Cecilia Leite Oliveira
Diretora

Reginaldo de Araújo Silva
Coordenação de Administração - COADM

Gustavo Saldanha
Coordenação de Ensino e Pesquisa, Ciência e Tecnologia da Informação - COEPE

José Luis dos Santos Nascimento
Coordenação de Planejamento, Acompanhamento e Avaliação - COPAV

Anderson Itaborahy
Coordenação-Geral de Pesquisa e Desenvolvimento de Novos Produtos - CGNP

Bianca Amaro de Melo
Coordenação-Geral de Pesquisa e Manutenção de Produtos Consolidados - CGPC

Tiago Emmanuel Nunes Braga
Coordenação-Geral de Tecnologias de Informação e Informática - CGTI

Milton Shintaku
Coordenação de Tecnologias para Informação - COTEC

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES
Instituto Brasileiro de Informação em Ciência e Tecnologia

MINISTÉRIO DA SAÚDE
Agência Nacional de Vigilância Sanitária

GUIA SOBRE A CONSTRUÇÃO DE TESAUROS

Milton Shintaku
Deise Maria Marcos Sabbag
Marcelle Costal
Raissa da Veiga Menêses

Brasília, DF
Instituto Brasileiro de Informação em Ciência e Tecnologia
Agência Nacional de Vigilância Sanitária
2021

EQUIPE TÉCNICA

Diretora do Instituto Brasileiro de Informação em Ciência e Tecnologia
Cecilia Leite Oliveira

Coordenador-Geral de Tecnologias de Informação e Informática – CGTI
Tiago Emmanuel Nunes Braga

Coordenador do Projeto
Milton Shintaku

Autores
Milton Shintaku
Deise Maria Marcos Sabbag
Marcelle Costal
Raíssa da Veiga de Menêses

Revisão de texto
Rafael Teixeira de Souza

Diagramação e projeto gráfico
Victor Ramos Silva

Normalização
Marcelle Costal
Raíssa da Veiga de Menêses

S556g Shintaku, Milton
Guia sobre a construção de tesauros / Milton Shintaku, Deise Maria Marcos Sabbag, Marcelle Costal, Raíssa da Veiga de Menêses. - Brasília: Ibict, 2021.

43 p.: il. color.

ISBN 978-65-89701-39-2
DOI 10.22477/9786589701392

1. Tesauros. 2. Sistema de organização do conhecimento. 3. Vocabulário controlado. 4. Software livre. I. Shintaku, Milton. II. Sabbag, Deise Maria Marcos. III. Costal, Marcelle. IV. Menêses, Raíssa da Veiga de. V. Instituto Brasileiro de Informação em Ciência e Tecnologia. VI. Título.

CDU 027(036):005.94

Ficha catalográfica elaborada por Marcelle Costal de Castro dos Santos RJ-00616/20

Esta produção é um produto do Projeto Estudos para implementação de Repositório Institucional na Biblioteca Terezinha Ayres Costa.
Ref. Processo SEI N° 01302.000479/2018-80 (Processo de Contratação)
Ref. Processo SEI N° 01302.000437/2020-63 (Processo de Execução)
Ref. FUNDEP - 28139

SUMÁRIO

1 APRESENTAÇÃO	06
2 INTRODUÇÃO	08
3 LINGUAGEM DE ESPECIALIDADE	10
4 CONCEITO	12
5 TERMO	15
6 TESAURO	18
7 MÉTODOS DE CONSTRUÇÃO DO TESAURO	19
8 FLUXO PROPOSTO PARA CONSTRUÇÃO DE TESAUROS	22
8.1 SELEÇÃO DE TERMOS	25
8.2 VERIFICAÇÃO DA SUA REPRESENTATIVIDADE E ESPECIFICIDADE	26
8.3 CONCEITUAÇÃO	27
8.3.1 NOTAS	28
8.3.2 ASPECTOS GRAMATICAIS	29
8.4 RELAÇÕES ENTRE TERMOS	31
8.4.1 RELAÇÃO HIERÁRQUICA	31
8.4.2 RELAÇÃO ASSOCIATIVA	33
8.4.3 RELAÇÃO DE EQUIVALÊNCIA	34
8.5 AVALIAÇÃO	35
8.6 REGISTRO, USO, DISSEMINAÇÃO E ATUALIZAÇÃO	38
9 CONSIDERAÇÕES FINAIS	40
REFERÊNCIAS	42

1 APRESENTAÇÃO

Este guia é resultado de estudos promovidos pelo projeto de pesquisa firmado entre a Agência Nacional de Vigilância Sanitária (ANVISA) e o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), voltado ao desenvolvimento de estudos sobre sistemas informatizados para acervo digital com vistas a serem implantados em sua biblioteca. Para tanto, entre as metas acordadas no projeto está o apoio ao desenvolvimento de sistema informatizado destinado a um Tesauros para a ANVISA, que pode ser utilizado como base de autoridade para outros sistemas.

Assim, este guia tem por objetivo amparar teoricamente a construção de Tesauros, responsabilidade da equipe da ANVISA, de forma a apoiar a padronização de termos e conceitos para a representação e recuperação da informação e do conhecimento no âmbito da ANVISA.

O escopo da proposta tem como ponto focal a Biblioteca Terezinha Ayres, da ANVISA, caracterizada como uma unidade de informação especializada em saúde pública, de acesso livre para servidores e usuários externos, tendo, em seu acervo, obras físicas e digitais direcionadas ao campo da vigilância sanitária e áreas correlatas. A coleção possui grande relevância social, econômica, política e pública formada por livros, trabalhos acadêmicos, obras de referência, periódicos, materiais especiais, produção intelectual da ANVISA (como artigos, trabalhos de conclusão de curso, dissertações, teses e outros produtos desenvolvidos pelos servidores em suas pesquisas acadêmicas) e publicações editadas pela própria Agência (ANVISA, 2021).

O Guia de apoio do desenvolvimento do Tesauros da ANVISA tem a finalidade de oferecer subsídios teóricos, metodológicos e práticos para o desenvolvimento do design do Tesauro. Dessa forma, contribui para a construção de uma ferramenta voltada à representação e recuperação da informação em saúde pública, auxiliando na efetivação do objetivo do projeto, que é o desenvolvimento de um sistema de informação para a disseminação da memória técnica da Agência.

Para tanto, este guia segue, entre outras, a norma *International Organization of Standardization*, criada pela *International Organization of Standardization* - ISO 25964-1 (2011): *Information and documentation - Thesauri and interoperability with other vocabularies - part 1*, visto que aporta a maioria das teorias sobre construção de Tesauros. Segue, também, as orientações apresentadas pelo *American National Standards Institute - National Information Standard Organization* - ANSI/NISO Z39-19 (2010) - *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Da mesma forma, utiliza a NBR 12.676 (1992) - Métodos para Análise de Documentos, Determinação de seus Assuntos, seleção de Termos para Indexação, da Associação Brasileira de Normas Técnicas (ABNT).

O princípio subjacente na construção de um Tesauro é a orientação para que o indexador e o autor intelectual escolham o mesmo termo para o mesmo conceito, garantindo que os documentos relevantes sejam recuperados. Ou seja, a norma estabelece sua orientação tendo como princípio a concepção de análise orientada para o conteúdo, mas não se deve esquecer da análise orientada para a demanda (usuário), bem como a demanda da instituição, por isso a importância da criação da política de indexação.

2 INTRODUÇÃO

Um dos paradigmas do projeto da modernidade é a busca pela especialidade que leva à fragmentação e compartimentalização dos conhecimentos. Nessa perspectiva, a partir de dados considerados mais sensíveis, ou particulares, busca-se a compreensão de verdades gerais. Nesse movimento, são criadas cada vez mais áreas de atuação profissional, para dar conta dessas especificidades, mesmo que estas possam ser contempladas por áreas maiores.

Dessa forma, pode-se classificar atividades de maneira mais ampla ou refinada. Há ainda casos de atividades transversais, em que atuam em mais de uma área, com relações múltiplas, dependendo da forma que se estabelece.

Para estruturar atividades, tópicos de estudo, formas de vidas e tantos temas, pode-se fazer uso de taxonomias, que podem ser mais amplas ou mais específicas, dependendo do objetivo. Assim, pode-se criar estruturas baseadas em objetivos específicos, como, por exemplo, a criação de taxonomias das atividades humanas por profissões ou formação acadêmica, em que ocorrem certas divergências, mas devem atender ao seu propósito. Desse modo, "pedreiro" aparece na taxonomia de "profissões", mas não na de "formação acadêmica", enquanto "médico" está presente nas duas.

Assim, uma taxonomia é um conjunto de termos organizados hierarquicamente, de forma a representar algo, conforme um objetivo. Evidentemente, toda taxonomia deve estar baseada em estudos para ser concisa e aceita

pela comunidade usuária. Uma das taxonomias mais conhecidas é a dos seres vivos, utilizada na biologia, criada no século XVIII, mas que ainda sofre ajustes, quando necessário. Outras taxonomias são mais particulares e atendem a um país, a comunidades ou mesmo a instituições, como as Áreas de Conhecimento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), com atuação somente no Brasil, pela área acadêmica.

Entretanto, em alguns casos, taxonomias requerem a adição de explicação simples dos termos que a compõem, para torná-los mais compreensíveis a todos, a exemplo do caso em que uma taxonomia pode ser utilizada por vários tipos de profissionais ou mesmo para ser apresentada ao público em geral. Dessa forma, possibilita o uso de termos corretos (preferidos) pelos profissionais ou usuário em geral, eliminando ambiguidades. No entanto, não se deve confundir Tesauros com glossários, ainda que ambos representem o conhecimento da especialidade.

Nesse contexto, os Tesauros situam-se nas chamadas linguagens de especialidades, utilizadas em áreas ou profissões específicas ou mesmo no registro de informações técnicas e científicas. São construídos seguindo termos aceitos pelos usuários, com definições simples, apenas para apresentá-los e possibilitar o uso. Possuem relações hierárquicas, tornando possível ao usuário ver termos mais gerais ou mais específicos, respeitando a área, as relações de equivalência e as relações associativas, elaboradas por meio do uso da linguagem de especialidade.

3 LINGUAGEM DE ESPECIALIDADE

Com o desenvolvimento das atividades humanas foi preciso criar termos que não gerassem ambiguidades, portanto, com representação precisa, formando os chamados jargões ou terminologia. Assim, na linguagem geral pode-se lançar mão de um determinado termo, mas, nas atividades profissionais ou na área do conhecimento, emprega-se o termo aceito na linguagem de especialidade, mesmo que a linguagem de especialidade seja um subconjunto de uma língua, utilizada por uma comunidade distinta.

A título de exemplo, na linguagem geral, muitas vezes as pessoas utilizam o termo "animal" para se referir a mamíferos, mesmo sabendo que insetos, peixes e anfíbios também são animais. Entretanto, os profissionais da área são precisos quanto à linguagem utilizada para se comunicar com os seus pares ao fazerem uso do termo correto, pertencente à linguagem de especialidade.

A linguagem de especialidade pertence a uma comunidade discursiva que a usa de modo consensual, como as línguas, que estão em processo contínuo de evolução, e possuem termos sendo inseridos, alterados, com novos conceitos, ou mesmo termos caindo em desuso, devido a novas descobertas, novos processos ou métodos. Da mesma

forma, empréstimos, neologismos e estrangeirismo, fenômenos comuns nas línguas, constantemente ocorrem na linguagem de especialidade, devido ao intercâmbio de conhecimento existente entre as pessoas da comunidade e o compartilhamento de informação.

Nesse sentido, o emprego da linguagem de especialidade possibilita o uso preciso da informação, evitando ambiguidades na comunicação entre os membros da comunidade. Pode também ser aplicada de forma didática na comunicação com a população, apresentando conhecimentos da área, principalmente na educação, nos livros didáticos. Assim, o uso das linguagens de especialidades é um instrumento útil na comunicação, independente do canal utilizado.

A linguagem de especialidade é constituída de termos que representam conceitos dentro de uma área, mesmo que para outra área tenha outra definição, sendo dotado de uma relação pragmática forte. "Protocolo", por exemplo, é um termo comum a comunidades de informática, administração, relações internacionais e medicina, porém com conceitos distintos, ainda que preserve semelhanças semânticas. Desse modo, em muitos casos, o uso de um termo será aplicado de acordo com os conceitos e a construção do campo semântico da área de atuação.

A complexidade da linguagem de especialidade não se apresenta somente na constituição de seus termos, mas nas possíveis dimensões, podendo ser de uma especialidade, área de atuação, região ou mesmo um grupo. Por isso a importância de instrumentos que registram os termos da linguagem de especialidade como forma de disseminação do conhecimento, podendo ser, inclusive, desenvolvido para determinado público. Ademais, esse registro possibilita a construção de sistemas de organização do conhecimento que tornam possível a organização e recuperação da informação, bem como a reflexão e problematização das questões históricas e culturais de sua formação.

4 CONCEITO

O processo de conceituação possibilita que possamos compreender os objetos que estão ao nosso redor mesmo sem que este seja diferenciado ou nomeado.

Os objetos são categorizados em classes, que correspondem a unidades de conhecimento chamadas de conceitos. Estes são representados na linguagem natural de várias formas: símbolos, códigos ou fórmulas, ícones, imagens, diagramas, gráficos, sinais, letras, classificação, linguagem de especialidade de uma área (por exemplo a área médica, a química, a física, a biologia etc.). Esta cartilha tratará dos conceitos que representam a terminologia em um campo especializado, ou seja, o conceito representará um termo.

O conceito faz a ligação entre o objeto e sua definição ou designação. Por isso, para a produção de uma linguagem de especialidade, uma terminologia, é necessária a compreensão da conceituação que sustenta o conhecimento na área específica do assunto.

Os conceitos podem ser gerais ou individuais. Aqueles não estão presos no tempo e no espaço, correspondendo a objetos gerais; estes estão presos no “aqui e agora”, no tempo e no espaço. Os conceitos gerais expressam objetos gerais; os conceitos individuais expressam objetos individuais.

O conceito geral corresponde a conjuntos de dois ou mais objetos que possuem propriedades comuns, a exemplo de: fundo de mercado imobiliário; esporte; música.

O conceito individual corresponde a um único objeto, ou uma composição única de entidades, por exemplo: Internet; Organização Mundial de Saúde (OMS); ANVISA.

CONCEITOS INDIVIDUAIS	CONCEITOS GERAIS
A USP	As Universidades
A Partida de São Paulo e Grêmio no dia 14 de setembro de 2014	As partidas de futebol
O Descobrimento do Brasil no ano de 1500	As descobertas marítimas

Quadro 1 - Conceitos Individuais e Gerais
Fonte: Elaborado pelos autores.

A linguagem natural possibilita a construção dos enunciados dos conceitos individuais e gerais. Portanto, o enunciado deve fazer referência a alguns dos elementos do conceito para ser verdadeiro.

CONCEITOS INDIVIDUAIS	CONCEITOS GERAIS
IBICT	INSTITUIÇÃO
ENUNCIADOS • uma instituição • localizada no Rio de Janeiro / Brasília • relacionada com a coordenação dos sistemas de informação no Brasil	ENUNCIADOS • constituído por pessoas • trabalham para um fim • localizada em determinado lugar • durante um tempo e em um espaço
A soma total dos enunciados verdadeiros sobre o IBICT fornece o conceito do mesmo	O conjunto de enunciados constitui o conceito de Instituição

Quadro 2 - Enunciados
Fonte: Elaborado pelos autores.

A linguagem natural possibilita a construção dos enunciados dos conceitos individuais e gerais. Portanto, o enunciado deve fazer referência a alguns dos elementos do conceito para ser verdadeiro.

NÍVEL	INDIVIDUAIS	GERAIS
Objetos	Objetos individuais	Objetos gerais
Conceitos	Conceitos individuais	Conceitos gerais
Sinais verbais	Nomes individuais	Nomes gerais
Sinais não-verbais	Sinais individuais	Sinais gerais

Quadro 3 - Formação dos Conceitos
Fonte: Elaborado pelos autores.

Importante: Todo enunciado possui elementos do objeto, isto é, atributos que são características que vão se repetir nas relações entre os conceitos, sejam: relações lógicas, relações hierárquicas, relações partitivas, relações de oposição ou relações funcionais.

5 TERMO

A língua é um componente importante da linguagem, constituída de um código verbal característico para a comunicação humana cotidiana. O código verbal da língua escrita é dinâmico e complexo, com elementos e variações, formado por diversos fenômenos e expressões linguísticas, como a ambiguidade, que pode levar a problemas na interpretação de enunciados, enunciados falsos, dificultando a comunicação.

Nas linguagens de especialidades, principalmente as ligadas à ciência e tecnologia, isso não é plausível. Nessas linguagens, o conceito é representado por meio de um termo preferido dentro de uma comunidade discursiva. O objetivo de atribuir um termo a um conceito é mantê-lo dentro de um mesmo campo semântico, ou seja, um TERMO deve comunicar um CONCEITO para que exista o controle da monossemia (monorreferencialidade) e um CONCEITO só seja representado por um TERMO (univocidade).

O termo tem origem no latim *terminus*, que significa final, ponto de conclusão, tanto que faz parte da família de outras palavras, como terminal, término, terminar. Termo representa o conceito preciso de um objeto, fenômeno, de algo aceito por uma comunidade. Dessa maneira, o uso do termo pela comunidade atribui o seu significado, o conceito aceito em suas múltiplas dimensões por essa comunidade.

Cabe, no entanto, destacar que termo não pode ser confundido com vocábulo, visto que pode se configurar como um conjunto de palavras que remetem a um conceito distinto. Geralmente, no português, os termos possuem uma

base, na qual pode-se agregar elementos que refinam o seu conceito. “Protocolo”, por exemplo, é um termo que ocorre em várias linguagens de especialidades, mas “protocolo de comunicação” é um termo da área de informática e “protocolo de atendimento” pertence à área de prestação de serviços. Em outras palavras, os complementos fornecem a especificidade necessária. Mesmo em uma área de conhecimento, a agregação pode refinar os termos. Assim, “doença” é um termo geral, “doença transmissível” restringe o termo geral a tipos específicos, “ doença transmissível por vetor” restringe ainda mais, observando que todos os termos são da mesma área do conhecimento.

DOENÇA
Doença Transmissível
Doença Transmissível por Vetor

Como parte integrante das atividades de uma comunidade, os termos podem sofrer fenômenos linguísticos, a exemplo das variações. No entanto, a comunidade geralmente seleciona uma opção para não gerar inconvenientes. Assim, nas linguagens de especialidades a presença de sinonímias não é comum, visto que sempre existe o termo mais utilizado, ou como apresentado nas linguagens de indexação, o termo autorizado.

A criação de novos termos costumeiramente se dá por meio de empréstimos de outra área, estrangeirismos ou acessando o fundo lexical da língua. Muitos termos da ciência originam-se do Latim, assim como da tecnologia do Grego, mas atualmente são originários do Inglês, mesmo que tenham origem greco-latina, como no caso do termo “delete” que foi importado da língua inglesa para a informática, mas que tem origem na palavra latina “*delere*”, apagar. O termo é pragmático, tanto que, pode-se criar termos com o fundo lexical, como no caso do termo "Síndrome da Imunodeficiência Adquirida”, com sigla SIDA, mas que, na prática, manteve o termo inglês *Acquired Immune Deficiency Syndrome* (AIDS)”.

Para facilitar a disseminação e o uso de termos, pode-se registrá-los em Sistemas de Representação da Informação, como os Tesauros, que contribuem para a disseminação da terminologia das áreas, facilitando o entendimento dos termos por todos, membros ou não da comunidade usuária, e, da mesma forma, apoiando as atividades de bibliotecários, jornalistas entre outros profissionais com atividades ligadas à informação.

Em um Tesauro, o termo preferível (termo descritor) é o termo autorizado a compor o Tesauro e o termo não-preferível (não-descritor ou termo equivalente) é um termo não autorizado, que denota uma variação do termo, utilizada como remissiva e não como termo autorizado. Por sua vez, o termo candidato é um termo assinalado pela equipe de especialistas, ou pelo usuário, sugerido para atualizar o Tesauro e, após verificação, pode se tornar termo autorizado, termo não preferível ou ser excluído do Tesauro. O termo qualificador é utilizado para diferenciar termos que possuem a mesma grafia ou especificar

o descritor.

Ex.: CAPITAL (ECONOMIA)
CAPITAL (ESTADO)

Os termos podem também estabelecer relações hierárquicas do geral para o específico, ou por associação.

6 TESAURO

O Tesauro é um vocabulário controlado, com base em uma estrutura semântica e funcional, classificado de acordo com as ideias que representa, a partir de um domínio ou área de conhecimento específicos. Para o estabelecimento dos termos e conceitos, o controle de vocabulário é essencial, evitando a ocorrência de fenômenos linguísticos, como: ambiguidade, plurissignificação, polissemia e homonímia, entre outros.

Em um vocabulário controlado, a estrutura relacional dos termos que representam os conceitos deve ser apresentada em ordem alfabética e hierarquicamente, para facilitar seu uso pelo indexador e pelos usuários do sistema, ou seja, facilitar o acesso ao conteúdo e seu posterior uso. Para que a informação seja visível, acessível e disponibilizada, é necessário que seja organizada nos sistemas de recuperação da informação (SRI), que, de modo abrangente, podem ser compreendidos como sistemas que irão proporcionar a interação, ou melhor, a interface, entre os mais diversos recursos informacionais existentes em uma coleção (impressa ou digital) e seus múltiplos usuários, possibilitando a busca e a recuperação dos documentos por meio de um índice. A sua estrutura e o seu vocabulário devem ser decididos na fase anterior, ou seja no planejamento, antes dos termos serem coletados.

7 MÉTODOS DE CONSTRUÇÃO DE TESAURO

Para a construção de Tesauro existem quatro abordagens indicadas na norma ANSI/NISO Z39.19 (2010), voltada à seleção de termos: abordagem de comitê, abordagem empírica, combinação de métodos e assistência da máquina.

A primeira, a *abordagem de comitê* é aquela em que especialistas de um determinado domínio ou área do conhecimento a que se destina o vocabulário controlado elaboram uma lista de termos chave e indicam a relação entre eles, amparados por especialistas em vocabulário controlado. Essas listas de termos podem ser retiradas de várias fontes ou submetidas aos usuários e comunidades envolvidas. Essa abordagem possui dois métodos: o descendente (*top down*) e o ascendente (*bottom up*).

a) Método descendente: os termos mais genéricos são identificados primeiro e então os termos mais específicos são selecionados para atingir o nível desejado de especificidade. As estruturas hierárquicas e relações são estabelecidas à medida que o trabalho evolui;

b) Método ascendente: ocorre quando as listas de termos são derivadas de um *corpus* de objetos de

de conteúdo e devem então ser incorporadas em um vocabulário controlado. Tal como no método descendente, as estruturas e relações hierárquicas necessárias são criadas à medida que o trabalho avança, mas o método ascendente inicia-se com os termos que têm uma abrangência mais específica e segue para os termos mais genéricos.

Se um novo Tesauro for construído, a abordagem "descendente" é a indicada. No caso de um Tesauro já publicado, a abordagem "ascendente" é mais usada para o acréscimo de novos termos necessários à cobertura de novos conceitos.

A segunda, a abordagem empírica (empirical approach), é baseada em dois métodos:

a) Método dedutivo: os termos são extraídos de objetos de conteúdo (por humanos ou computadores) durante um estágio preliminar de indexação, mas sem pensar no controle do vocabulário, até que um número suficiente de termos tenha sido coletado. Todos os termos são revisados por um grupo de especialistas, preferencialmente, composto de especialistas em informação e no assunto tratado. Identifica-se primeiramente os termos que representam as classes mais amplas, a fim de alocar os termos remanescentes naquelas classes segundo suas relações lógicas, de modo que as hierarquias possam ser estabelecidas do *mais genérico* para o *mais específico*. O controle de vocabulário é desenvolvido no estágio em que as hierarquias e outras relações são estabelecidas;

b) Método indutivo: novos termos são selecionados para potencial inclusão no Tesauro na forma como são encontrados em objetos de conteúdo. O controle de vocabulário é aplicado desde o início. Se o vocabulário em construção tiver algum tipo de arranjo hierárquico, cada termo admitido é designado como um membro de uma ou mais classes genéricas, que são construídas sobre uma base a*d hoc*, em um estágio anterior. O vocabulário controlado é, portanto, estabelecido sobre uma base de termos que vai do *mais específico* para o mais *genérico*. A construção de vocabulário controlado é considerada uma operação contínua. A assistência de especialistas no assunto é altamente recomendada, sendo que podem atuar como membros de uma comissão ou comitê editorial ou formal.

A terceira abordagem, a combinação de métodos, é a aplicação de uma dessas abordagens em um estágio ou outro durante a construção de um vocabulário controlado. A título de exemplo, hierarquias e outras relações entre termos que foram inicialmente estabelecidos indutivamente podem, mais tarde, ser examinadas a partir de um ponto de vista dedutivo. Ambas as técnicas são essencialmente empíricas.

É certo que a construção de um Tesauro é uma atividade intelectual, mas pode receber o auxílio da assistência tecnológica. Assim, a quarta abordagem, a assistência da máquina

(machine assistance), é empregada em tarefas de identificação de termos, tais como:

a) Identificação de termos candidatos: identificados automaticamente, a partir de texto legível por máquina (por exemplo: títulos e/ou resumos). Os termos que não são registrados podem ser considerados candidatos à inclusão;

b) Registro da frequência da indicação de termo: em sistemas de indexação computadorizados, a frequência com a qual um termo é usado em indexação pode ser registrada automaticamente. Termos com resultados altos ou baixos podem ser considerados candidatos à modificação ou exclusão;

c) Registro de termos a partir das consultas do usuário: os termos encontrados nas consultas de usuário que não combinam com um ou mais termos de entrada também podem ser considerados para inclusão especialmente quando um dado termo ocorre em múltiplas consultas. Para assegurar a privacidade, os usuários não devem ser identificados quando se rastreia e coleta consultas de usuário. Termos que ocorrem em linguagem natural e são formas comuns de descrever um conceito são incluídos no Tesauro como termos de entrada. Um termo de entrada é uma palavra ou frase que é um nome comumente usado ou rótulo para um conceito, mas não é o termo preferido, selecionado para representar o conceito.

Após a escolha do método a ser empregado, sugere-se que a equipe elabore um fluxo para a construção, que irá definir os parâmetros que irão compor o Tesauro.

8 FLUXO PROPOSTO PARA CONSTRUÇÃO DE TESAUROS

A construção de Tesauros requer cuidado, pois precisa ser amparada por padrões para acuidade dos vocabulários. Para tal, os seguintes documentos orientadores da obra foram:

- ISO 25964-1 (2011): norma que cobre todos os aspectos para o desenvolvimento e manutenção de Tesauros monolíngues e multilíngues destinados à recuperação de informação;
- NBR 12676 (1992): norma que cobre os métodos para análise de documentos, determinação de seus assuntos e seleção de termos de indexação;
- ANSI/NISO Z39.19 (2010): apresenta as diretrizes para a construção, para o formato e gestão de vocabulários controlados monolíngues.

Planejar um Tesauro é tarefa que deve ser realizada em conjunto por uma equipe multidisciplinar, de forma cuidadosa, tendo em vista a sua consistência. A fase de

planejamento de um Tesauro tem como princípio norteador o conhecimento sobre as normas e um exercício prévio de classificação. Isto pode ser estruturado a partir de um estudo de frequência de termos, cunhados no arcabouço técnico e teórico de determinada temática, domínio ou área do conhecimento. Tal exercício é composto por uma categorização para delimitar as classes de maior abrangência e determinar os termos que irão compor o Tesauro (CAMPOS; GOMES; MOTA, 2004).

É importante identificar no planejamento o público-alvo a quem o Tesauro se destina, uma vez que ele é especializado por natureza. Nesse sentido, tanto a seleção dos termos quanto o nível de especificidade é diferente, pois depende do tipo de usuário, não existindo a possibilidade de elaborar um Tesauro geral que dê conta de cobrir diferentes temáticas.

Com o intuito de fornecer uma base para orientar a construção do Tesauro da ANVISA, este guia propõe um fluxo que cobre os principais assuntos abordados nas normas. O fluxo de construção de Tesauros não tem o objetivo de ser exaustivo, uma vez que para aprofundar a dúvida é preciso recorrer às normas. Tampouco precisa ser seguido à risca, de forma engessada, pois, de acordo com as necessidades identificadas na fase de planejamento, a equipe pode mudar o caminho sugerido.

Desse modo, a Figura 1 apresenta uma síntese do fluxo sugerido na proposta, delimitado nas seções subsequentes.



Figura 1 - Fluxo proposto para construção de Tesauros da Anvisa
Fonte: Elaborado pelos autores.

Com base no fluxo proposto, e tendo em vista a tomada de decisão sobre quais áreas de assunto irão compor o Tesauro, a seleção de termos é a primeira atividade para iniciar a sua construção.

8.1 SELEÇÃO DE TERMOS

A seleção dos termos de um Tesauro ocorre no corpus documental, constituído por documentos técnicos, Linguagens Documentárias pertencentes a um Sistema de Organização do Conhecimento (SOC) ou por uma revisão de literatura da área em bases de dados referenciais. Logo, consiste no levantamento de um vocabulário a ser organizado com o indicativo das fontes utilizadas. Para iniciar o levantamento dos termos, um registro preliminar pode ser realizado por meio de listas terminológicas, mas é sugerido que eles sejam indexados no próprio software para controle e coordenação entre as equipes.

A indexação dos termos é realizada a partir de uma leitura flutuante (leitura documentária), que irá direcionar a análise de conteúdo dos documentos selecionados e a designação do seu respectivo assunto. Desse modo, a partir do assunto será delimitado o conceito e, por conseguinte, a tradução do conceito em uma linguagem de indexação, que é designada por um termo.

A escolha do termo deve estar pautada no grau de especificidade definido pela equipe e no uso de descritores correntes e reconhecidos pelos pares. Tal definição, porém, precisa estar expressa na Política de Indexação da Instituição. Todos os termos selecionados para compor o Tesauro devem estar dentro da linguagem de especialidade, ou seja, do domínio da área a ser coberta. Caso haja a recorrência de algum termo novo, indica-se a consulta a especialistas e a instrumentos de referência. O auxílio de instrumentos como os códigos de classificação, dicionários, glossários e outros Tesauros são importantes nessa fase.

Na estruturação de um Tesauro, um descritor é o sinônimo de termo e representa um conceito autorizado para indexar e recuperar um determinado assunto. Em um mesmo documento pode existir uma variação na nomenclatura dos termos e, portanto, o termo não descritor é não equivalente. Embora o termo não descritor expresse o mesmo conceito que o preferido, não pode ser indexado como entrada principal.

O termo autorizado a compor o Tesauro deve expressar um conceito único, evitando a presença de ambiguidades e variações como sinonímias, homonímias, homografias, regionalismos etc. A sua normalização segue o padrão da língua portuguesa, que é o masculino singular, mas precisa verificar as exceções para os casos em que os termos são apresentados unicamente na forma feminina. Assim, a forma do termo pode ser um substantivo, um adjetivo substantivado ou um verbo substantivado:

ALIMENTO (substantivo)
COLIDÊNCIA (adjetivo substantivado)
ARTICULAÇÃO (verbo substantivado)

Termos com mais de uma palavra, formando expressões, devem representar uma ideia indivisível, mesmo que compostos por partes não atendidas pelos Tesauros. Por exemplo:

ADMINISTRAÇÃO EM SAÚDE PÚBLICA (Termo ou Unidade Terminológica que expressa complexidade.)

No entanto, no processo de normalização gramatical dos termos autorizados, sempre que possível os substantivos devem constar no singular, para facilitar o processo de recuperação. Contudo, em determinados casos, o uso do plural é necessário, por configurar um conjunto de elementos, como a seguir:

DADOS PESSOAIS (termo no plural)
RELAÇÕES INTERNACIONAIS (termo no plural)

No caso dos substantivos próprios, como os nomes geográficos e organizações governamentais, indica-se criar listas auxiliares e não integrá-las no corpo do Tesauro.

Por fim, a qualidade da indexação depende da impessoalidade na escolha terminológica, no amparo às fontes e especialistas da área, das necessidades organizacionais e da adequação à linguagem dos usuários.

8.2 VERIFICAÇÃO DA SUA REPRESENTATIVIDADE E ESPECIFICIDADE

Antes dos termos serem admitidos em um Tesauro, eles devem ser validados de acordo com os princípios de seleção e determinação, estando conforme os critérios de escopo, forma e escolha dos termos adotados pela equipe.

Os termos precisam estar em conformidade com a linguagem utilizada pelos especialistas no assunto por meio das fontes autorizadas.

Se a área de abrangência do Tesauro compreender mais de um domínio ou área do conhecimento, pode ser necessário criar microtesauros, como vocabulário mais específico e ligado ao vocabulário do Tesauro Geral, que tem um critério menos específico sobre a temática escolhida pela Organização. Tal estratégia é importante para definir os níveis de

especificidade, facilitar o gerenciamento dos termos e controlar a duplicação de termos e sinônimos.

8.3 CONCEITUAÇÃO

A conceituação de termos é uma das etapas da Representação Temática da Informação e da Organização do Conhecimento. A seleção de termos compilados da análise documental são formas verbais que sintetizam e representam um conceito (DAHLBERG, 1978). Por sua vez, a conceituação tem como objetivo compreender e identificar os conceitos que irão compor uma linguagem documentária. Para a construção de Tesauros, a conceituação é uma atividade intelectual, que busca compreender os significados implícitos por meio da soma das características essenciais ou optativas da análise do objeto, restringido a uma área de conhecimento, ou mais áreas, no caso de ser um Tesauro multidisciplinar.

Ou seja, conceito é uma unidade de pensamento e pode ser expressado de forma simples, como “agrotóxico”, ou complexas, como “Sistema Nacional de Vigilância Sanitária”.

De acordo com a ISO 25964-1 (2011), um conceito representado por termo em Tesauro pode pertencer a categorias que compartilham as mesmas características, tais como:

a) Coisas e as suas partes físicas:
MEMBROS DO CORPO
PÁSSAROS

b) Materiais:
BORRACHA
ADESIVO

c) Atividades e processos:
FERTILIZAÇÃO
ACONDICIONAMENTO

d) Eventos e ocorrências:
PANDEMIAS
GUERRAS CIVIS

e) Entidades abstratas, propriedades de pessoas, coisas, materiais ou ações:
PERSONALIDADE
ELASTICIDADE
RAPIDEZ

f) Disciplinas ou campos temáticos
QUÍMICA ORGÂNICA
FARMACOLOGIA

g) Unidades de medida:
QUILÔMETROS
HERTZ

h) Tipos de pessoas e organizações
ORGANIZAÇÕES GOVERNAMENTAIS
PESSOA COM DEFICIÊNCIA

As entidades únicas, expressas por nomes próprios (pessoas, documentos e objetos), podem ser reunidas em um Tesauro, desde que sejam utilizadas apenas com a finalidade de registro. Nos casos em que não for assunto principal dos documentos, devem estar como um vocabulário auxiliar e não na estrutura do Tesauro em si. Nomes próprios possuem categorias, tais como:

a) Lugares:
RIO DE JANEIRO (RJ)
REINO UNIDO

b) Objetos específicos, recursos topográficos e outras entidades:
CARTA MAGNA
MONALISA

c) Pessoas físicas, cargos organizacionais e entidades coletivas:
GHEBREYESUS, TEDROS ADHANOM
DIRETOR GERAL DA ORGANIZAÇÃO MUNDIAL DA SAÚDE
ORGANIZAÇÃO MUNDIAL DA SAÚDE

8.3.1 NOTAS

Em situações em que um termo não expresse com clareza a totalidade de um conceito, sendo muito restrito ou amplo, é sugerido o uso de notas de escopo. Embora ela não faça parte da estrutura conceitual do termo, deve ser usada para esclarecer os limites de um conceito. A nota de escopo pode ser usada para fornecer outros detalhes sobre o uso de termos para a comunicação entre indexadores ou para o pesquisador. Ela não precisa apresentar uma definição completa, mas deve esclarecer o uso pretendido de um termo no Tesauro. Outros tipos de nota podem ser transmitidos como nota de definição, que

apresenta a definição do termo e a nota histórica que orienta o usuário na mudança de uma terminologia.

a) Nota de escopo:
ILUMINURA
NE: Inclui a decoração ornamental e as ilustrações em manuscritos, bem como alguns livros impressos, se feitos à mão.

b) Nota de definição:
BOAS PRÁTICAS CLÍNICAS
DEF: Padrão para o planejamento, a condução, a realização, o monitoramento, a auditoria, o registro, a análise e o relato de ensaios clínicos que fornecem garantia de que os dados e os resultados relatados têm credibilidade e precisão, e que os direitos, a integridade e o sigilo dos participantes do ensaio clínico estão protegidos, de acordo com as diretrizes de BPC dispostas no Documento das Américas e no Manual de Boas Práticas Clínicas da Conferência Internacional de Harmonização.

c) Nota histórica:
MICROONDAS
NH: Conceito introduzido em 1985.

8.3.2 ASPECTOS GRAMATICAIS

Alguns aspectos gramaticais devem ser observados na escolha e definição dos termos do Tesauro. De acordo com a ISO 25964-1 (2011), algumas diretrizes a serem seguidas são:

- Os termos de um Tesauro devem ser, preferencialmente, **substantivos** ou **frases nominais**;

- O uso de **adjetivos** sozinhos pode causar problemas na recuperação da informação, portanto, eles devem ser evitados. Apesar disso, podem ser utilizados em frases nominais;

- Os **advérbios** não devem ser usados em termos de Tesauros;

- Os **verbos** não devem ser usados isoladamente, portanto, atividades devem ser representadas por substantivos ou frases nominais.

Indexar termo é um processo de atribuir termos preferidos para descrever os conceitos de determinado objeto, e isto cobre sistemas e procedimentos intelectuais de seleção e organização da informação (ANSI/NISO Z39.19, 2010). Também envolve a inclusão de

- **Eliminação de ambiguidades, homonímia (polissemia)**: A ambiguidade é um processo que ocorre por meio do uso da linguagem natural e expressa o uso de termos similares, mas que possuem significados diferentes.

Podem ser homográficos (que têm a mesma escrita) ou homófonos (que têm a mesma fonética).

Ex.: **Homógrafos**
MERCÚRIO (METAL)
MERCÚRIO (PLANETA)
MERCÚRIO (MITOLOGIA)

Homófonos
CESSÃO (DE BENS)
SEÇÃO (DE REFERÊNCIA)
SESSÃO (DE CINEMA)

Em um Tesauro, a ambiguidade provocada pelos homófonos e homógrafos é diferenciada por qualificadores.

Ex.: ADJUDICAÇÃO (LICITAÇÃO) [qualificador do termo]

- **Controle de sinônimos (sinonímia)**: ocorre quando um conceito pode ser representado por diferentes palavras. É preciso tomar uma decisão sobre qual dentre as palavras utilizar (a de uso mais recorrente) e colocar os outros termos semelhantes como termos não preferidos.

Ex.: PENITENCIÁRIA
NÃO USE Cadeia
NÃO USE Presídio
NÃO USE Sistema penitenciário

- **Antonímia**: são termos que apresentam significados opostos que, no caso de um Tesauro, podem ser prefixos ou palavras. Um termo antônimo é indexado em um Tesauro, não em relação a um termo contrário, mas quando o seu significado possui informações adicionais sobre o assunto.

Ex.: PROBIDADE ADMINISTRATIVA / IMPROBIDADE ADMINISTRA

Outros tipos de princípios para expressar de forma mais apropriada os conceitos incluem o estabelecimento de relações semânticas entre termos.

8.4 RELAÇÕES ENTRE TERMOS

O Tesauro pode manifestar vários tipos de relações semânticas, com a finalidade de articular os conceitos indexados. Tem a função de auxiliar os indexadores a determinar o nível de especificidade de um conceito. Como visto anteriormente, um Tesauro pode estabelecer relações de igualdade por meio da equivalência de termos semelhantes, mas também expressa relações hierárquicas e associativas.

8.4.1 RELAÇÃO HIERÁRQUICA

A relação hierárquica pode ser estabelecida entre conceitos quando o escopo de um deles é parte do outro, determinando graus ou níveis de superordenação e subordinação. O conceito superordenado representa uma classe ou totalidade e o subordinado são membros ou partes do todo. O Termo Geral (TG) representa o termo superordenado e o Termo Específico (TE) representa o termo subordinado. Tanto o termo mais amplo quanto o mais específico devem representar uma mesma categoria inerente, seja ela uma coisa, uma ação, uma propriedade etc. (ISO 25964-1, 2011).

Ex.: (TG) ANIMAIS
(TE) MAMÍFEROS

(TG) METAIS
(TE) OURO

Os exemplos "metais" e "ouro" representam materiais e podem, portanto, estar hierarquicamente relacionados.

O que não pode ser feito: "metais" (uma classe de materiais) e "fundição" (uma ação) representam diferentes tipos de conceito e, portanto, não podem ser relacionados hierarquicamente.

Por sua vez, uma relação hierárquica pode ser de três situações (ANSI/NISO Z39.19, 2010):

a) Relação genérica: ligação entre uma classe e categoria com os seus membros ou espécies, ou ainda, na relação entre todos e alguns.

Ex.: Alguns membros da classe "pássaros" são conhecidos como "papagaios", e todos os "papagaios", por definição, e independentemente do contexto, são considerados

"pássaros". O teste geralmente garante que um termo como "papagaios" não esteja subordinado a uma classe como "animais de estimação", pois nem todos os papagaios são animais de estimação (ISO 25964-1, 2011, p. 59), ao menos que papagaio corresponda a um Tesauro especializado em animais domésticos.

b) Relação hierárquica (relação de todo parte): relação em que a parte de uma entidade ou sistema pertence exclusivamente a um todo particular. São abreviados por TGP (Termo Genérico Partitivo) e TEP (Termo Específico Partitivo). Pode ser aplicada a quatro classes principais:

- Sistemas e órgãos do corpo:
  TGP: SISTEMA CARDIOVASCULAR
  TEP: VEIAS SANGUÍNEAS
  TEP: ARTÉRIA
  TEP: VEIAS

- Localizações geográficas:
  TGP: BRASIL
  TEP: DISTRITO FEDERAL
  TEP: ASA SUL
  TEP: ASA NORTE

- Disciplinas ou campos de discurso:
  TGP: CIÊNCIAS
  TEP: BIOLOGIA
  TEP: BOTÂNICA
  TEP: ZOOLOGIA

- Estruturas sociais e hierárquicas:
  TGP: IGREJA
  TEP: IGREJA CATÓLICA
  TEP: CLERO
  TEP: LEIGOS

c) Relação de instância: identifica o vínculo entre uma categoria geral de coisas ou eventos, expressa por um substantivo comum, e uma instância individual dessa categoria, geralmente um nome próprio.

TGI: REGIÕES MONTANHOSAS
TEI: ALPES HIMALAIAS

No exemplo acima, os Alpes e o Himalaia são atribuídos a posições subordinadas em uma hierarquia, embora não sejam nem tipos nem partes de regiões montanhosas, mas representam exemplos ou instâncias específicas.

Alguns conceitos podem estabelecer relacionamentos poli-hierárquicos, pertencendo a mais de uma categoria. No exemplo a seguir, o termo PIANO está subordinado a dois termos genéricos.

Ex1.: TG: INSTRUMENTO DE CORDA
TE: PIANO

Ex2.: TG: INSTRUMENTO DE PERCUSSÃO
TE: PIANO

Outros termos presentes no Tesauro em uma estrutura hierárquica do tipo "árvore" apenas agregam assuntos. São chamados "rótulos de nó", sendo utilizados para mostrar os princípios de divisão entre um conjunto de "termos irmãos" (termos que agregam termos mais amplos), que têm a mesma função dos termos gerais, mas que não devem ser usados como termos de indexação. Eles são diferenciados dos termos autorizados por meio de colchetes angulares, itálico etc.

Ex.: <GESTÃO>
TG: GESTÃO DA REGULAMENTAÇÃO
TE: ANÁLISE DE RECURSOS

8.4.2 RELAÇÃO ASSOCIATIVA

Outro tipo de relacionamento é o associativo. Não estabelece relações hierárquicas e nem de equivalência, mas relações semânticas ou conceitualmente associadas, que necessitam estarem expostas no Tesauro. Recebem o nome de Termos Relacionados (RT), podendo ser vinculados aos termos de mesma hierarquia, ou a termos pertencentes a diferentes hierarquias.

Ex.1: CÉLULAS
TR: CITOLOGIA
Ex.2: CITOLOGIA
TR: CÉLULAS

Os termos "irmãos", ou seja, aqueles que compartilham o mesmo termo mais amplo, não necessitam de associação, pois são mutuamente exclusivos. Rosa e Azaléia são termos específicos de Flores, e não precisam estabelecer um inter-relacionamento, uma vez que o

significado de uma ou outra não se sobrepõe.

8.4.3 RELAÇÃO DE EQUIVALÊNCIA

No processo de construção de um Tesauro é comum encontrar mais de um termo, na mesma língua, para representar determinado conceito. Quando isso ocorre, apenas uma forma deve ser designada como **termo preferido**, enquanto as demais formas devem ser inseridas como **termos não preferidos**. A partir disso, são estabelecidas as relações de equivalência: **USE** e **USADO PARA** (UP). A escolha da melhor forma a ser utilizada deve considerar aspectos como o assunto do Tesauro e aos usuários a que ele se destina. Dentro do Tesauro, essa relação é apresentada da seguinte forma:

Ex.: MANDIOCA **UP** Aipim
Aipim **USE** MANDIOCA

No exemplo acima, o **termo preferido**, ou seja, aquele que é autorizado e deverá ser usado como descritor, é MANDIOCA. Dessa forma, o termo não preferido, aquele que não poderá ser utilizado como descritor, é AIPIM.

As relações **USE** e **UP** podem ocorrer por diferentes motivos, entre eles:

- **Sinônimos**: termos que podem ser considerados iguais em determinado idioma. Neste caso, os termos preferidos devem ser escolhidos de acordo com as necessidades dos usuários.

- **Termos específicos agrupados em um termo geral**: em alguns casos, a depender do tamanho e abrangência do Tesauro, um termo mais amplo pode ser escolhido como termo preferido, tendo seus termos específicos como não preferidos. Isso pode ser feito, também, para diminuir a quantidade de termos preferidos dentro de um Tesauro.

Ex.: PEDRA PRECIOSA **UP** Esmeralda
UP Diamante
UP Turmalina
Esmeralda **USE** PEDRA PRECIOSA
Diamante **USE** PEDRA PRECIOSA
Turmalina **USE** PEDRA PRECIOSA

- **Ortografia**: deve-se utilizar a ortografia mais aceita dos termos. Se existirem variações ortográficas que sejam comumente aceitas, elas devem ser inseridas como termos não preferidos, com referência ao termo preferido. Termos com erros ortográficos não podem ser usados como preferidos. Apesar disso, quando um erro ortográfico é muito comum, ele pode ser inserido como termo não preferido.

- **Estrangeirismo e suas traduções**: palavras emprestadas de outras línguas (estrangeirismo) devem ser usadas como termo preferido quando já incorporadas ao vocabulário e cotidiano da comunidade.

Ex.: MILKSHAKE
OUTDOOR

Além disso, algumas palavras emprestadas de outras línguas já foram incorporadas à língua portuguesa a ponto de receberem traduções e adaptações. Quando isso acontece, o termo preferido deve ser aquele mais amplamente utilizado, enquanto os demais devem ser referenciados como termo não preferido.

Ex.: SHAMPOO (forma em inglês) → XAMPU (forma aportuguesada)
BALLET (forma em francês) → BALÉ (forma aportuguesada)

- **Abreviações**: a forma por extenso de um determinado conceito deve, em geral, ser o termo preferido, com referência à sigla como termo não preferido.

Ex.: SUPERIOR TRIBUNAL DE JUSTIÇA **UP** STJ
STJ **USE** SUPERIOR TRIBUNAL DE JUSTIÇA

Porém, em alguns casos, as siglas podem ser usadas como termo preferido quando são amplamente estabelecidas, especialmente quando a forma por extenso raramente é utilizada.

Ex.: UNESCO **UP** Organização das Nações Unidas para a Educação, a Ciência e a Cultura.
Organização das Nações Unidas para a Educação, a Ciência e a Cultura **USE** UNESCO

- **Nome de Lugares**: nomes de países e regiões geográficas podem variar entre os idiomas. De maneira geral, deve-se utilizar, como termo preferido, a forma mais popular entre a comunidade de usuários do Tesauro.

Ex.: NOVA IORQUE
NEW YORK

8.5 AVALIAÇÃO

De acordo com a ANSI/NISO Z39.19 (2010), o teste e a avaliação são atividades indispensáveis para um Tesauro a fim de determinar se está sendo usado para descrever os objetos de conteúdo, bem como fornecer resultados de pesquisa adequados para a

maioria dos usuários, correspondendo às suas expectativas. Para que o Tesauro retorne com precisão ao termo de busca dos usuários, é preciso que ele esteja enriquecido com termos usuais e que seus semelhantes estejam como termos não preferíveis, indicando qual o termo autorizado.

Para a avaliação do Tesauro, apontam-se alguns métodos de teste:

a) Avaliação heurística: conta com o apoio de um especialista ou grupo de especialistas para avaliar o Tesauro, podendo ser uma avaliação informal e qualitativa (o especialista compartilha as suas reações e opiniões) ou formal e quantitativa (os especialistas atribuem pontuações de acordo com uma lista compartilhada de critérios).

b) Modelagem de Afinidade: a equipe convida um grupo expressivo de usuários para pontuar termos equivalentes e relacionados de acordo com o nível de similaridade.

c) Teste de Usabilidade: avaliação do sistema de informação sobre a eficácia do Tesauro orientada pela literatura nas áreas de interação humano-computador e arquitetura de informação no tocante à usabilidade e métodos de avaliação.

Somados à metodologia, podem ser formulados critérios para orientar a avaliação do Tesauro. Na lista abaixo constam alguns critérios.

Sinônimos	• O Tesauro contém termos que são sinônimos ou equivalentes sem indicar a relação? • Estão sendo atribuídos aos termos autorizados? • O que foi necessário para a construção dos termos sinônimos? Verifique as fontes usadas.
Análise conceitual	• Quão bem o Tesauro reflete a análise conceitual (por exemplo, sequência de conceitos em algum nível hierárquico?)
Notas	• As notas históricas são fornecidas quando necessário? • Estão incluídas as notas de escopo?

Facetas	• Todas as facetas necessárias estão incluídas?
Termos	• Os termos estão atualizados de acordo com a utilização aceita? (periodicidade dos termos)
Forma dos termos	• A forma do termo é consistente? • A forma de termo segue o uso comum?
Tesauros acessíveis eletronicamente	• As referências cruzadas são hiperlinks ativos? • Existe um índice de termos on-line ou pesquisa por palavra-chave? • O banco de dados de vocabulário controlado é pesquisável? Como funciona a pesquisa? Quais campos são pesquisados?
Procedimentos para o desenvolvimento do Tesauro	• Verifique os procedimentos usados no desenvolvimento de vocabulário controlado. • Examine a estrutura e o conteúdo do vocabulário controlado. • Use o conhecimento da estrutura do vocabulário controlado para análise da estrutura e consistência interna. • Verifique outros Tesauros, enciclopédias, dicionários ou outras fontes oficiais.

	• Verifique como funciona o vocabulário controlado. Tente usá-lo para solicitações de pesquisa e objetos de conteúdo de indexação. Nesse caso, o conhecimento das áreas temáticas é necessário para a equipe. Experimentos de indexação em que várias pessoas indexam os mesmos itens podem ser muito úteis; divergências podem apontar problemas no vocabulário controlado.

Quadro 4 - Critérios de avaliação de Tesauros
Fonte: Elaborado pelos autores, com base na ANSI/NISO Z39.19 (2010).

Esses critérios fornecem uma base para estruturar o processo de avaliação do Tesauro. Outros podem ser acrescentados, de acordo com as observações apontadas pela equipe.

8.6 REGISTRO, USO, DISSEMINAÇÃO E ATUALIZAÇÃO

O registro de termos é importante para o controle do vocabulário e permite manter a memória dos termos que compõem o Tesauro. Ele deve ser feito no momento da admissão do termo. Para os casos de perda de termos no sistema, é considerável manter um registro individual, em ficha, para cada um, permitindo o acesso e a recuperação, bem como a atualização dos termos caso estejam em desuso. No registro, identifica-se quais foram as autoridades consultadas, as fontes utilizadas e a data de inclusão no sistema. Ele reúne todo o histórico do termo: termos sinônimos, termos específicos e termos relacionados.

Com relação ao uso a que se destina o Tesauro, observa-se que as instituições possuem perfis próprios, que podem ser evidenciados por meio dos materiais de seus acervos. De acordo com Dodebei (2002), o Tesauro é um instrumento que pode ser usado para indexar e recuperar informações. Nesta perspectiva, o Tesauro, por ser uma linguagem documentária, pode ser utilizado para o tratamento temático das informações institucionais, uma vez que é um instrumento de representação de uma área ou domínio. Assim, ele pode classificar e traduzir conceitos com coerência, ampliando o entendimento do conceito por meio da semântica.

A rede de relações presente no Tesauro amplia a sua disseminação, pois permite que um termo ultrapasse a sua forma e conteúdo, proporcionando ao usuário maior revocação sobre o que ele realmente procura (DODEBEI, 2002). Assim, por ser um Sistema de Organização do do Conhecimento, o Tesauro como instrumento de comunicação possibilita a organização e a gestão do conhecimento dentro da Agência, acelerando a sua disseminação.

De acordo com a norma ISO 25964-1 (2011), um Tesauro possui diferentes formas de disseminação, de modo que a forma a ser adotada pela equipe deve ser escolhida no momento do planejamento. Este pode ser integrado a um sistema eletrônico, ser uma publicação eletrônica (CD ROM, Website ou internet) ou distribuído em cópia impressa de forma limitada, já que é um produto interno. No entanto, cada vez mais o Tesauro assume o formato eletrônico, mais especificamente em formato XML. Ele precisa ter um módulo de exportação com o objetivo de ser facilitada a utilização para outros fins e por outras pessoas.

O Tesauro naturalmente precisa acompanhar as mudanças terminológicas, uma vez que existem termos em que o seu uso é descontinuado, ou seja, ao longo de sua vida útil, ele necessita de uma previsão de alterações (ISO 25964-1, 2011). Portanto, a terminologia tem que ser revisada contando com uma periodicidade, de acordo com a política de atualização. Um formulário pode ser encaminhado à equipe para identificar sugestões de novos conceitos, termos preferidos, termos não preferidos, correção de termos existentes e exclusão de termos. Existem softwares que agregam mecanismos que permitem adicionar termos candidatos no Tesauro de modo privado, para que em conjunto se possa definir se o termo será validado ou não.

9 CONSIDERAÇÕES FINAIS

A oferta de informação pela web tem provocado significativas mudanças na forma de atuação do público em geral, pois coloca as pessoas em contato com temas que podem não ser comuns, apresentando especificidades pouco conhecidas. Com isso, requer, em grande parte dos casos, ferramentas de disseminação de terminologias, como os Tesauros, para apoiar a compreensão correta das informações apresentadas.

A elaboração de um Tesauro deve atender a uma instituição, disciplina, atuação profissional etc., com a representação da linguagem de especialidade, com um conjunto de termos e conceitos apresentados de forma sistemática e organizada.

Logo, este guia procurou apresentar os aspectos teóricos da construção de vocabulários controlados para apoiar o registro de termos e o trabalho contínuo de atualização do Sistema de Tesauros da ANVISA, tendo como principal função apresentar à equipe da Agência um compilado sobre os principais assuntos expressos nas normas orientadoras para Tesauros monolíngues. Com isso, visa apoiar a construção de Tesauros da Agência, ofertando a teoria e a orientação prática na seleção, avaliação e registro dos termos e seus conceitos, no âmbito do projeto de pesquisa firmado entre a ANVISA e o Ibict.

Esta cartilha atende ao cumprimento de metas acordadas entre a Agência e a instituição, no desenvolvimento do projeto de pesquisa, ofertando documentação técnica, resultados de estudos,

apresentando os métodos para análise de documentos, determinação de seus assuntos e seleção de termos de indexação, desenvolvimento e manutenção de Tesauros, bem como as diretrizes para a construção, para formato e gestão de vocabulários controlados.

Por fim, espera-se que esta cartilha possa ser utilizada como apoio a outras instituições e órgãos de governo na construção de seus Tesauros, visto que os resultados de uma pesquisa aplicada tem por finalidade cooperar para o desenvolvimento de soluções aos problemas reais. Contribui-se, desse modo, com a oferta de documentação técnica sobre a construção de Tesauros, mesmo sem o aprofundamento de um livro ou manual, mas com a praticidade de uma cartilha.

O Tesauro naturalmente precisa acompanhar as mudanças terminológicas, uma vez que existem termos em que o seu uso é descontinuado, ou seja, ao longo de sua vida útil, ele necessita de uma previsão de alterações (ISO 25964-1, 2011). Portanto, a terminologia tem que ser revisada contando com uma periodicidade, de acordo com a política de atualização. Um formulário pode ser encaminhado à equipe para identificar sugestões de novos conceitos, termos preferidos, termos não preferidos, correção de termos existentes e exclusão de termos. Existem softwares que agregam mecanismos que permitem adicionar termos candidatos no Tesauro de modo privado, para que em conjunto se possa definir se o termo será validado ou não.

REFERÊNCIAS

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. **Biblioteca**. Brasília, DF: ANVISA, 10 de maio de 2021. Disponível em: http://antigo.anvisa.gov.br/biblioteca. Acesso em: 10 maio. 2021.

AMERICAN NATIONAL STANDARDS INSTITUTE. ANSI/NISO Z39.19-2005. **Guidelines for the construction, format and management of monolingual thesauri**. Bethesda: ANSI, 2005. Disponível em: https://groups.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf. Acesso em: 25 jan. 2021.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 12676. **Métodos para análise de documentos**: determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro: ABNT, 1992.

CAMPOS, Maria Luiza Almeida; GOMES, Hagar Espanha; MOTTA, Dilza Fonseca. **Critérios de avaliação de tesauros**. Rio de Janeiro: Conexão Rio, 2004. Disponível em: http://www.conexaorio.com/biti/tesauro/index.htm. Acesso em: 8 mar. 2021.

DAHLBERG, Ingetraut. Referent-oriented, analytical concept theory for interconcept. **Intern. Classificat.**, [S. l.], v. 5, n. 3, p. 142-151, 1978.

DODEBEI, Vera Lúcia Doyle Louzada de Mattos. **Tesauro**: linguagem de representação da memória documentária. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002.

INTERNATIONAL ORGANIZATION OF STANDARDIZATION. ISO 25964-1. **Information and documentation**: thesauri and interoperability with other vocabularies - part 1: Thesauri for information retrieval. Geneva: ISO, 2011. Disponível em: https://www.iso.org/standard/53657.html. Acesso em: 14 jan. 2021

ISBN: 978-65-89701-39-2
9 786589 701392